

Bounds for stochastic processes on product index spaces

Witold Bednorz ^{*†‡§}

February 1, 2016

Abstract

In this paper we discuss the question how to bound supremum of a stochastic process with the index set of a product type. There is a tempting idea to approach the question by the analysis of the process on each of the marginal index spaces separately. However it turns out that we also need to study suitable partitions of the whole index space. We show what can be done in this direction and how to use the method to reprove some known results. In particular we observe that all known applications of the Bernoulli Theorem can be obtained in this way, moreover we use the shattering dimension to slightly extend the application to VC classes. We also show some application to the regularity of paths for processes which take values in vector spaces. Finally we give a short proof of the Mendelson-Paouris result on sums of squares for empirical processes.

1 Introduction

In this paper I denotes a countable set and $(F, \|\cdot\|)$ a separable Banach space. Consider the class \mathcal{A} of subsets of I . We say that the class \mathcal{A} *satisfies the maximal inequality* if for any symmetric independent random variables $X_i, i \in I$ taking values in F the following inequality holds

$$\mathbf{E} \sup_{A \in \mathcal{A}} \left\| \sum_{i \in A} X_i \right\| \leq K \mathbf{E} \left\| \sum_{i \in I} X_i \right\|, \quad (1.1)$$

where K depends on \mathcal{A} only. We stress that in this paper K will be used to denote constants that appear in the formulation of our results and may depend on their structure. We use c, C, L, M to denote absolute constants which may change their values from line to line by numerical factors. Also we write \sim to express that two quantities are comparable up to a universal constant. This will help us to reduce the notation applied in this paper. It is an easy observation to see that (1.1) is equivalent to

$$\mathbf{E} \sup_{A \in \mathcal{A}} \left\| \sum_{i \in A} v_i \varepsilon_i \right\| \leq K \mathbf{E} \left\| \sum_{i \in I} v_i \varepsilon_i \right\|, \quad (1.2)$$

^{*}**Subject classification:** Primary 60G15; Secondary 60G17

[†]**Keywords and phrases:** VC classes, shattering dimension, stochastic inequalities

[‡]Institute of Mathematics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

[§]Research partially supported by MNiSW Grant N N201 608740 and Mobility Plus

where $(v_i)_{i \in I}$, consists of vectors in F and $(\varepsilon_i)_{i \in I}$ is a Bernoulli sequence, i.e. a sequence of independent r.v.'s such that $\mathbf{P}(\varepsilon_i = \pm 1) = \frac{1}{2}$.

To understand what is the right characterization of such classes \mathcal{A} we have to recall the notion of VC dimension. We say that \mathcal{A} has VC dimension d if there exists a set $B \subset I$, $|B| = d$ such that $|\{B \cap A : A \in \mathcal{A}\}| = 2^d$ but for all $B \subset I$, $|B| > d$, $|\{B \cap A : A \in \mathcal{A}\}| < 2^{d+1}$. It means that \mathcal{A} shatters some set B of cardinality d but do not shatter any set of cardinality $d + 1$. The result which has been proved in [1] as a corollary of the Bernoulli Theorem states that finite VC dimension is the necessary and sufficient condition for the class \mathcal{A} to have the property (1.1). Since our paper strongly refers to the Bernoulli Theorem we recall its formulation. We start from mentioning Talagrand's result for Gaussian's processes. In order to find two-sided bounds for supremum of the process $G(t) = \sum_{i \in I} t_i g_i$, where $t \in T \subset \ell^2(I)$ and $(g_i)_{i \in I}$ is a Gaussian sequence, i.e. a sequence of independent standard Gaussian r.v.'s we need Talagrand's $\gamma_2(T)$ numbers, cf. Definition 2.2.19 in [15] or (3.1) below. By the well known Theorem 2.4.1 in [15] we have

$$\mathbf{E} \sup_{t \in T} G(t) \sim \gamma_2(T). \quad (1.3)$$

The Bernoulli Theorem, i.e. Theorem 1.1 in [1], concerns a similar question for processes of random signs.

Theorem 1.1 *Suppose that $T \subset \ell^2(I)$. Then*

$$\mathbf{E} \sup_{t \in T} \sum_{i \in I} t_i \varepsilon_i \sim \inf_{T \subset T_1 + T_2} \left(\sup_{t \in T_1} \|t\|_1 + \gamma_2(T_2) \right).$$

where the infimum is taken over all decompositions $T_1 + T_2 = \{t_1 + t_2 : t_1 \in T_1, t_2 \in T_2\}$ that contain the set T and $\|t\|_1 = \sum_{i \in I} |t_i|$.

Note that if $0 \in T$ then we can also require that $0 \in T_2$ in the above result. The consequence of Theorem 1.1 to our problem with maximal inequalities is as follows.

Theorem 1.2 *The class \mathcal{A} satisfies (1.1) with a finite constant K if and only if \mathcal{A} is a VC class of a finite dimension. Moreover the square root of the dimension is up to a universal constant comparable with the optimal value of K .*

Observe that part of the result is obvious. Namely one can easily show that if \mathcal{A} satisfies the maximal inequality then it is necessarily a VC class of a finite dimension. Indeed let $(\varepsilon_i)_{i \in I}$ be a Bernoulli sequence. Suppose that set $B \subset I$ is shattered. Let $x_i = 1$ for $i \in B$ and $x_i = 0$, $i \notin B$. Obviously

$$\mathbf{E} \left| \sum_{i \in I} x_i \varepsilon_i \right| = \mathbf{E} \left| \sum_{i \in B} \varepsilon_i \right| \leq \sqrt{|B|}$$

and on the other hand

$$\mathbf{E} \sup_{A \in \mathcal{A}} \left| \sum_{i \in A} x_i \varepsilon_i \right| = \mathbf{E} \sup_{A \in \mathcal{A}} \left| \sum_{i \in A \cap B} x_i \varepsilon_i \right| \geq \mathbf{E} \sum_{i \in B} \varepsilon_i 1_{\varepsilon_i = 1} = |B|/2.$$

Consequently if (1.1) holds then $K \geq \sqrt{|B|}/2$. Therefore (1.1) implies that cardinality of B must be smaller or equal $4K^2$.

Much more difficult is to prove the converse statement, i.e that for each VC class \mathcal{A} of dimension d the inequality (1.2) holds with K comparable with \sqrt{d} . In order to prove the result one has to first replace the basic formulation of the maximal inequality - (1.2) by its equivalent version

$$\mathbf{E} \sup_{A \in \mathcal{A}} \sup_{t \in T} \left| \sum_{i \in A} t_i \varepsilon_i \right| \leq K \mathbf{E} \sup_{t \in T} \left| \sum_{i \in I} t_i \varepsilon_i \right|, \quad (1.4)$$

where $(\varepsilon_i)_{i \in I}$ is a Bernoulli sequence and $0 \in T \subset \ell^2(I)$. Note that we use absolute values since part of our work concerns complex spaces. However it is important to mention that in the real case $\mathbf{E} \sup_{t \in T} \sum_{i \in I} t_i \varepsilon_i$ is comparable with $\mathbf{E} \sup_{t \in T} |\sum_{i \in I} t_i \varepsilon_i|$ if $0 \in T$ and therefore we often require in this paper that $0 \in T \subset \ell^2(I)$. Let us denote

$$b(T) = \mathbf{E} \sup_{t \in T} \left| \sum_{i \in I} t_i \varepsilon_i \right|, \quad g(T) = \mathbf{E} \sup_{t \in T} \left| \sum_{i \in I} t_i g_i \right|,$$

where $(\varepsilon_i)_{i \in I}$, $(g_i)_{i \in I}$ are respectively Bernoulli and Gaussian sequence. We recall that, what was known for a long time [5, 7], (1.4) holds when Bernoulli random variables are replaced by Gaussians, i.e.

$$\mathbf{E} \sup_{A \in \mathcal{A}} \sup_{t \in T} \left| \sum_{i \in A} g_i t_i \right| \leq C \sqrt{d} \mathbf{E} \sup_{t \in T} \left| \sum_{i \in I} t_i g_i \right| = C \sqrt{d} g(T), \quad (1.5)$$

for any $0 \in T \subset \ell^2(I)$. Due to Theorem 1.1 one can cover the set T by $T_1 + T_2$, where $0 \in T_2$ and

$$\max \left\{ \sup_{t \in T_1} \|t\|_1, g(T_2) \right\} \leq L b(T). \quad (1.6)$$

Therefore using (1.5) and (1.6)

$$\begin{aligned} & \mathbf{E} \sup_{A \in \mathcal{A}} \sup_{t \in T} \left| \sum_{i \in A} \varepsilon_i t_i \right| \\ & \leq \sup_{t \in T_1} \|t\|_1 + \mathbf{E} \sup_{A \in \mathcal{A}} \sup_{t \in T_2} \left| \sum_{i \in A} \varepsilon_i t_i \right| \\ & \leq \sup_{t \in T_1} \|t\|_1 + \sqrt{\frac{\pi}{2}} \mathbf{E} \sup_{A \in \mathcal{A}} \sup_{t \in T_2} \left| \sum_{i \in A} t_i \varepsilon_i \mathbf{E} |g_i| \right| \\ & \leq \sup_{t \in T_1} \|t\|_1 + \sqrt{\frac{\pi}{2}} \mathbf{E} \sup_{A \in \mathcal{A}} \sup_{t \in T_2} \left| \sum_{i \in A} t_i g_i \right| \\ & \leq \sup_{t \in T_1} \|t\|_1 + \sqrt{\frac{\pi}{2}} g(T_2) \leq CL \sqrt{d} b(T). \end{aligned}$$

This proves Theorem 1.2.

We show another example in which a similar approach works. Let G be a compact Abelian group and $(v_i)_{i \in I}$ a sequence of vectors taking values in F .

Let $\chi_i, i \in I$ be characters on G . The deep result of Fernique [4] is

$$\mathbf{E} \sup_{h \in G} \left\| \sum_{i \in I} v_i \chi_i(h) g_i \right\| \leq C \left(\mathbf{E} \left\| \sum_{i \in I} v_i g_i \right\| + \sup_{\|x^*\| \leq 1} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} x^*(v_i) \chi_i(h) g_i \right| \right).$$

This can be rewritten similarly as (1.5), i.e. for any $0 \in T \subset \ell^2(I)$ (which is a complex space in this case)

$$\mathbf{E} \sup_{h \in G} \sup_{t \in T} \left| \sum_{i \in I} t_i \chi_i(h) g_i \right| \leq C \left(g(T) + \sup_{t \in T} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) g_i \right| \right). \quad (1.7)$$

Once again the Bernoulli Theorem allows to prove a similar result for Bernoulli sequences. Namely by Theorem 1.1 we get the decomposition $T \subset T_1 + T_2$, $0 \in T_2$ such that

$$\max \left\{ \sup_{t \in T_1} \|t\|_1, g(T_2) \right\} \leq Lb(T). \quad (1.8)$$

Consequently using (1.7), (1.8) and $|\chi_i(h)| \leq 1$ we get

$$\begin{aligned} & \mathbf{E} \sup_{h \in G} \sup_{t \in T} \left| \sum_{i \in I} t_i \chi_i(h) \varepsilon_i \right| \\ & \leq \sup_{t \in T_1} \|t\|_1 + \mathbf{E} \sup_{h \in G} \sup_{t \in T_2} \left| \sum_{i \in A} \varepsilon_i t_i \chi_i(h) \right| \\ & \leq \sup_{t \in T_1} \|t\|_1 + \sqrt{\frac{\pi}{2}} \mathbf{E} \sup_{h \in G} \sup_{t \in T_2} \left| \sum_{i \in I} t_i \chi_i(h) g_i \right| \\ & \leq \sup_{t \in T_1} \|t\|_1 + C \left(g(T_2) + \sup_{t \in T_2} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) g_i \right| \right) \\ & \leq CL \left(b(T) + \sup_{t \in T_2} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) g_i \right| \right). \end{aligned} \quad (1.10)$$

The final step is the Marcus-Pisier estimate [12] (see Theorem 3.2.12 in [15])

$$\sup_{t \in T_2} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) g_i \right| \leq M \sup_{t \in T_2} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) \varepsilon_i \right|. \quad (1.11)$$

Note that (1.11) is deeply based on the translational invariance of the distance

$$d_t(g, h) = \left(\sum_{i \in I} |t_i|^2 |\chi_i(g) - \chi_i(h)|^2 \right)^{\frac{1}{2}} \quad g, h \in G. \quad (1.12)$$

Since we may assume that $T_2 \subset T - T_1$ we get

$$\begin{aligned}
& \sup_{t \in T_2} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) \varepsilon_i \right| \\
& \leq \sup_{t \in T} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) \varepsilon_i \right| + \sup_{t \in T_1} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) \varepsilon_i \right| \\
& \leq \sup_{t \in T} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) \varepsilon_i \right| + Lb(T). \tag{1.13}
\end{aligned}$$

Combining (1.10) with (1.11) and (1.13) we get the following result.

Theorem 1.3 *Suppose that $0 \in T \subset \ell^2(I)$. For any compact group G and a collection of vectors $v_i \in F$ in a complex Banach space $(F, \|\cdot\|)$ and characters χ_i on G the following holds*

$$\mathbf{E} \sup_{h \in G} \sup_{t \in T} \left| \sum_{i \in I} t_i \chi_i(h) \varepsilon_i \right| \leq K \left(b(T) + \sup_{t \in T} \mathbf{E} \sup_{h \in G} \left| \sum_{i \in I} t_i \chi_i(h) \varepsilon_i \right| \right).$$

The aim of this note is to explore questions described above in a unified language. Simply we consider random processes $X(u, t)$, $(u, t) \in U \times T$ with values in \mathbb{R} or \mathbb{C} , which means we study stochastic processes defined on product index sets. In particular we cover all canonical processes in this way. Indeed, suppose that $U \subset \mathbb{R}^I$ or \mathbb{C}^I and $T \subset \mathbb{R}^I$ or \mathbb{C}^I are such that for any $a \in U$ and $t \in T$ we have that $\sum_{i \in I} |u_i t_i|^2 < \infty$. Then for any family of independent random variables X_i such that $\mathbf{E} X_i = 0$, $\mathbf{E} |X_i|^2 = 1$,

$$X(u, t) = \sum_{i \in I} u_i t_i X_i, \quad u \in U, t \in T$$

is a well defined process. As we have mentioned our main classes of examples concern Gaussian canonical processes where $X_i = g_i$, $i \in I$ are standard normal variables or Bernoulli canonical processes where $X_i = \varepsilon_i$, $i \in I$ are random signs. In particular we aim to find bounds for $\mathbf{E} \sup_{u \in U} \left\| \sum_{i \in I} u_i v_i \varepsilon_i \right\|$, where $v_i \in F$, $i \in I$, formulated in terms of $\mathbf{E} \left\| \sum_{i \in I} v_i \varepsilon_i \right\|$. One of the results we state is the application of the shattering dimension introduced by Mendelson and Vershynin [14], which enable us to generalize Theorem 1.2. In this way we deduce that under mild conditions imposed on $\mathcal{A} \subset \mathbb{R}^I$ we have

$$\mathbf{E} \sup_{u \in U} \left\| \sum_{i \in I} u_i v_i X_i \right\| \leq K \mathbf{E} \left\| \sum_{i \in I} v_i X_i \right\|$$

for any independent symmetric r.v.'s X_i , $i \in I$. We show how to apply the result in the analysis of convex bodies and their volume in high dimensional spaces. On the other hand we can use the approach to study processes $X(t) = (X_i(t))_{i \in I}$, $t \in [0, 1]$ which take values in \mathbb{R}^I or \mathbb{C}^I . For example to check whether paths $t \rightarrow X(t)$ belong to ℓ^2 we should consider

$$X(u, t) = \sum_{i \in I} u_i X_i(t), \quad u = (u_i)_{i \in I} \in U, t \in [0, 1],$$

where U is the unit ball in $\ell^2(I)$, i.e. $U = \{u \in \mathbb{R}^I : \sum_{i \in I} |u_i|^2 \leq 1\}$. The finiteness of $\|X(t)\|_2 < \infty$ is equivalent to the finiteness of $\sup_{u \in U} |X(u, t)|$. Similarly we can treat a well known question in the theory of empirical processes. Suppose that $(\mathcal{E}, \mathcal{B})$ is a measurable space and \mathcal{F} a countable family of measurable real functions on \mathcal{E} . Let X_1, X_2, \dots, X_N be independent random variables which take values in $(\mathcal{E}, \mathcal{B})$, we may define

$$X(u, f) = \sum_{i=1}^N u_i f(X_i), \quad u = (u_i)_{i=1}^N \in U, f \in \mathcal{F},$$

where $U = B_N(0, 1) = \{u \in \mathbb{R}^N : \sum_{i=1}^N |u_i|^2 \leq 1\}$. Then it is clear that

$$\sup_{u \in U} |X(u, f)|^2 = \sum_{i=1}^N |f(X_i)|^2, \quad \text{for all } f \in \mathcal{F}.$$

In the last section we give a short proof of Mendelson-Paouris result [13] that gives an upper bound for $\mathbf{E} \sup_{u \in U} \sup_{f \in \mathcal{F}} |X(u, t)|$.

2 Upper bounds

For the sake of completeness we give an idea how to bound stochastic processes. The approach we present slightly extends former results of Latala [9, 10] and Mendelson-Paouris [13]. Suppose that $\mathbf{E}|X(t) - X(s)| < \infty$ for all $s, t \in T$. For each $s, t \in T$ and $n \geq 0$ we define $\bar{q}_n(s, t)$ as the smallest $q > 0$ such that

$$\begin{aligned} F_{q,n}(s, t) &= \mathbf{E} q^{-1} (|X(t) - X(s)| - q)_+ \\ &= \int_1^\infty \mathbf{P}(|X(t) - X(s)| > qt) dt \leq N_n^{-1}. \end{aligned} \quad (2.1)$$

We prove the following observation.

Lemma 2.1 *Function $\bar{q}_n(s, t)$, $s, t \in T$ is a distance on T , namely is symmetric, satisfies the triangle inequality and $\bar{q}_n(s, t) = 0$ if and only if $X(s) = X(t)$ a.s.*

Proof. Obviously $\bar{q}_n(s, t)$ is finite and symmetric $\bar{q}_n(s, t) = \bar{q}_n(t, s)$. To see that it equals 0 if and only if $\mathbf{P}(|X(t) - X(s)| > 0) > 0$ a.s. note that if $X(s) \neq X(t)$ then $\mathbf{E}|X(t) - X(s)| > 0$. The function $q \rightarrow F_{q,n}(s, t)$ is decreasing continuous and $F_{q,n}(s, t) \rightarrow \infty$ if $q \rightarrow 0$ and $\mathbf{E}|X(t) - X(s)| > 0$. Moreover $F_{q,n}(s, t)$ is strictly decreasing on the interval $\{q > 0 : F_{q,n}(s, t) > 0\}$ and consequently $\bar{q}(s, t)$ is the unique solution of $F_{q,n}(s, t) = N_n^{-1}$, namely

$$\mathbf{E}(\bar{q}_n(s, t))^{-1} (|X(t) - X(s)| - \bar{q}_n(s, t))_+ = N_n^{-1}.$$

Finally we show that \bar{q} satisfies the triangle inequality. Indeed for any $u, v, w \in T$ either $\bar{q}(u, v) = 0$ or $\bar{q}(v, w) = 0$ or $\bar{q}(u, w) = 0$ and the inequality is trivial or all the quantities are positive and then

$$F_{\bar{q}_n(u,v),n}(u, v) = F_{\bar{q}_n(v,w),n}(v, w) = F_{\bar{q}_n(u,w),n}(u, w) = N_n^{-1}.$$

It suffices to observe

$$\begin{aligned} & \frac{1}{\bar{q}_n(u, v) + \bar{q}_n(v, w)} \mathbf{E} (|X(u) - X(w)| - \bar{q}_n(u, v) - \bar{q}_n(w, v))_+ \\ & \leq \mathbf{E} \left(\frac{|X(u) - X(w)| + |X(w) - X(v)|}{\bar{q}_n(u, v) + \bar{q}_n(w, v)} - 1 \right)_+. \end{aligned}$$

The function $x \rightarrow (x - 1)_+$ is convex which implies that

$$(px + qy - 1)_+ \leq p(x - 1)_+ + q(y - 1)_+ \quad (2.2)$$

for $p, q \geq 0$, $p + q = 1$ and $x, y > 0$. We use (2.2) for

$$x = \frac{|X(u) - X(v)|}{\bar{q}_n(u, v)}, \quad y = \frac{|X(v) - X(w)|}{\bar{q}_n(v, w)}$$

and

$$p = \frac{\bar{q}_n(u, v)}{\bar{q}_n(u, v) + \bar{q}_n(w, v)}, \quad q = \frac{\bar{q}_n(v, w)}{\bar{q}_n(u, v) + \bar{q}_n(w, v)}.$$

Therefore

$$\begin{aligned} & \frac{1}{\bar{q}_n(u, v) + \bar{q}_n(v, w)} \mathbf{E} (|X(u) - X(w)| - \bar{q}_n(u, v) - \bar{q}_n(w, v))_+ \\ & \leq p \mathbf{E} \left(\frac{|X(u) - X(v)|}{\bar{q}_n(u, v)} - 1 \right)_+ + q \mathbf{E} \left(\frac{|X(v) - X(w)|}{\bar{q}_n(v, w)} - 1 \right)_+ \\ & \leq p N_n^{-1} + q N_n^{-1} = N_n^{-1} \end{aligned}$$

which by definition gives that

$$\bar{q}_n(u, v) + \bar{q}_n(v, w) \geq \bar{q}_n(v, w).$$

■

Obviously usually we do not need to use the optimal distances \bar{q}_n and replace the construction which is sufficient for our purposes. We say that a family of distances q_n , $n \geq 0$ on T is *admissible* if $q_n(s, t) \geq \bar{q}_n(s, t)$ and $q_{n+1}(s, t) \geq q_n(s, t)$. For example Latala [9, 10] and Mendelson-Paouris [13] used moments, namely if $\|X(t) - X(s)\|_p = (\mathbf{E}|X(t) - X(s)|^p)^{\frac{1}{p}} < \infty$, $p \geq 1$ then we may take $q_n(s, t) = 2\|X(t) - X(s)\|_{2^n}$. Indeed observe that

$$\mathbf{E} \left(\frac{|X(t) - X(s)|}{2\|X(t) - X(s)\|_{2^n}} - 1 \right)_+ \leq \frac{\mathbf{E}|X(t) - X(s)|^{2^n}}{(2\|X(t) - X(s)\|_{2^n})^{2^n}} \leq \frac{1}{N_n}.$$

Following Talagrand we say that a sequence of partitions \mathcal{A}_n , $n \geq 0$ of a set T is *admissible* if it is increasing, $\mathcal{A}_0 = \{T\}$ and $|\mathcal{A}_n| \leq N_n$. Let us also define admissible sequences of partitions $\mathcal{A} = (\mathcal{A}_n)_{n \geq 0}$ of the set T , which means nested sequences of partitions such that $\mathcal{A}_0 = \{T\}$ and $|\mathcal{A}_n| \leq N_n$. For each $A \in \mathcal{A}_n$ we define

$$\Delta_n(A) = \sup_{s, t \in A} q_n(s, t).$$

By $A_n(t)$ we call denote the element of \mathcal{A}_n that contains point t . For each $A \in \mathcal{A}_n$, $n \geq 0$ we define t_A as an arbitrary point in A . We may and will

assume that if $t_A \in B \in \mathcal{A}_{n+1}$, $B \subset A \in \mathcal{A}_n$ then $t_B = t_A$. Let $\pi_n(t) = t_{A_n(t)}$, then $\pi_0(t) = t_T$ is a fixed point in T . Let $T_n = \{\pi_n(t) : t \in T\}$ for $n \geq 0$. Clearly T_n , $n \geq 0$ are nested, namely $T_n \subset T_{n+1}$ for $n \geq 0$. For each stochastic process $X(t)$, $t \in T$ and $\tau \geq 0$ we may define

$$\gamma_X^\tau(T) = \inf_{\mathcal{A}} \sup_{t \in T} \sum_{n=0}^{\infty} \Delta_{n+\tau}(A_n(t)).$$

We prove that for $\tau \geq 2$, $\gamma_X^\tau(T)$ is a good upper bound for $\mathbf{E} \sup_{s,t \in T} |X(t) - X(s)|$.

Theorem 2.2 *For $\tau \geq 2$ the following inequality holds*

$$\mathbf{E} \sup_{s,t \in T} |X(t) - X(s)| \leq 4\gamma_X^\tau(T).$$

Proof. Note that

$$\begin{aligned} & |X(t) - X(\pi_0(t))| \\ & \leq \sum_{n=0}^{\infty} q_{n+\tau}(\pi_{n+1}(t), \pi_n(t)) \\ & + \sum_{n=0}^{\infty} (|X(\pi_{n+1}(t)) - X(\pi_n(t))| - q_{n+\tau}(\pi_{n+1}(t), \pi_n(t)))_+ \\ & \leq \sum_{n=0}^{\infty} \Delta_{n+\tau}(A_n(t)) + \sum_{n=0}^{\infty} \sum_{u \in T_n} \sum_{v \in A_n(u) \cap T_{n+1}} (|X(u) - X(v)| - q_{n+\tau}(u, v))_+. \end{aligned}$$

For any $\varepsilon > 0$ one can find nearly optimal admissible partition $(\mathcal{A}_n)_{n \geq 0}$ such that

$$\sup_{t \in T} \sum_{n=0}^{\infty} \Delta_{n+\tau}(A_n(t)) \leq (1 + \varepsilon) \gamma_X^\tau(T)$$

and therefore

$$\begin{aligned} & \mathbf{E} \sup_{t \in T} |X(t) - X(\pi_0(t))| \\ & \leq \gamma_X^\tau(T) + \varepsilon + \sum_{n=0}^{\infty} \sum_{u \in T_n} \sum_{v \in A_n(u) \cap T_{n+1}} \frac{q_{n+\tau}(u, v)}{N_{n+\tau}} \\ & \leq (1 + \varepsilon) \gamma_X^\tau(T) + \sum_{n=0}^{\infty} \sum_{u \in T_n} \Delta_{n+\tau}(A_n(u)) \frac{N_{n+1}}{N_{n+\tau}} \\ & \leq \gamma_X^\tau(T) \left(1 + \varepsilon + \sum_{n=0}^{\infty} \frac{N_n N_{n+1}}{N_{n+\tau}} \right) \\ & \leq (1 + \varepsilon) \gamma_X^\tau(T) + \gamma_X^\tau(T) \left(\frac{N_0 N_1}{N_2} + \sum_{n=1}^{\infty} \frac{1}{N_n} \right) \\ & \leq \gamma_X^\tau(T) \left(1 + \varepsilon + \sum_{n=0}^{\infty} 2^{-n-1} \right), \end{aligned}$$

where in the last line we used $N_n^2 = N_{n+1}$, $N_n \geq 2^{n+1}$ for $n \geq 1$ and $N_0 N_1 / N_2 = 1/4 < 1/2$. Since ε is arbitrary small we infer that $\mathbf{E} \sup_{t \in T} |X(t) - X(\pi_0(t))| \leq 2\gamma_X^\tau(T)$ and hence $\mathbf{E} \sup_{s,t} |X(t) - X(s)| \leq 4\gamma_X^\tau(T)$. ■

The basic question is how to construct admissible sequences of partitions. The simplest way to do this goes through entropy numbers. Recall that for a given distance ρ on T the quantity $N(T, \rho, \varepsilon)$ denotes the smallest number of balls of radius ε with respect to ρ necessary to cover the set T . Consequently for a given $\tau \geq 0$ we define entropy numbers as

$$e_n^\tau = \inf\{\varepsilon > 0 : N(T, q_{n+\tau}, \varepsilon) \leq N_n\}, \quad n \geq 0.$$

Having the construction ready we may easily produce an admissible sequence of partitions. On each level n there exists at most N_n sets of $q_{n+\tau}$ diameter 2ε that covers T . To obtain nested sequence of partitions we have to intersect all the sets constructed at levels $0, 1, \dots, n-1$. The partition \mathcal{A}_n has no more than $N_0 N_1 \dots N_{n-1} \leq N_n$ elements. Moreover for each set $A \in \mathcal{A}_n$ we have

$$\Delta_{n+\tau-1}(A) \leq 2e_{n-1}^\tau \quad \text{for } n \geq 1.$$

Obviously $\Delta_{\tau-1}(T) \leq \Delta_\tau(T) \leq 2e_0^\tau$. Let $\mathcal{E}_X^\tau(T) = \sum_{n=0}^\infty e_n^\tau$, then for any $\tau \geq 1$

$$\gamma_X^{\tau-1}(T) \leq 2e_0^\tau + 2 \sum_{n=1}^\infty e_{n-1}^\tau \leq 4\mathcal{E}_X^\tau(T).$$

and hence by Theorem 2.2 with $\tau \geq 3$

$$\mathbf{E} \sup_{s,t \in T} |X(t) - X(s)| \leq 16\mathcal{E}_X^\tau(T). \quad (2.3)$$

We turn to our main question of processes on product spaces. Consider $X(u, t)$, $u \in U$, $t \in T$ we define two different families of distances $q_{n,t}$ and $q_{n,u}$ admissible to control marginal processes respectively $u \rightarrow X(u, t)$ on U and $t \rightarrow X(u, t)$ on T .

First for a given $t \in T$, let $q_{n,t}$, $n \geq 0$ be a family of distances on U admissible for the process $u \rightarrow X(u, t)$ and let $e_{n,t}^\tau$, $n \geq 0$ be entropy numbers on U constructed for $q_{n,t}$, $n \geq 0$. By (2.3) we infer that for $\tau \geq 3$

$$\mathbf{E} \sup_{u,v \in U} |X(u, t) - X(v, t)| \leq 16 \sum_{n=0}^\infty e_{n,t}^\tau. \quad (2.4)$$

If we define $\mathcal{E}_{X,t}^\tau(U) = \sum_{n=0}^\infty e_{n,t}^\tau$ and $\mathcal{E}_{X,T}^\tau(U) = \sup_{t \in T} \mathcal{E}_{X,t}^\tau(U)$ then we may rewrite (2.4) as

$$\sup_{t \in T} \mathbf{E} \sup_{u,v \in U} |X(u, t) - X(v, t)| \leq 16\mathcal{E}_{X,T}^\tau(U),$$

and in this way the entropy numbers may be used to bound the family of processes $u \rightarrow X(u, t)$ where $t \in T$.

On the other hand for a given $u \in U$ let $q_{n,u}$, $n \geq 0$ be a family of distances on T admissible for $t \rightarrow X(u, t)$. Obviously $q_{n,U} = \sup_{u \in U} q_{n,u}$ is a good upper bound for all distances $q_{n,u}$. Let

$$\gamma_{X,U}^\tau(T) = \inf_{\mathcal{A}} \sup_{t \in T} \sum_{n=0}^\infty \Delta_{n+\tau,U}(A_n(t))$$

where the infimum is taken over all admissible partitions $\mathcal{A} = (\mathcal{A}_n)_{n \geq 0}$ of T and

$$\Delta_{n,U}(A) = \sup_{s,t \in A} q_{n,U}(s,t) = \sup_{s,t \in A} \sup_{u \in U} q_{n,u}(s,t).$$

Theorem 2.2 applied to distances $q_{n,U}$, $n \geq 0$ implies that

$$\sup_{u \in U} \mathbf{E} \sup_{s,t \in T} |X(u,t) - X(u,s)| \leq 4\gamma_{X,U}^\tau(T).$$

We prove that the described two quantities i.e. $\mathcal{E}_{X,T}^\tau(U)$ and $\gamma_{X,U}^\tau(T)$ suffice to control the process $X(u,t)$, $u \in U$, $t \in T$.

We state our main result which extends the idea described in Theorem 3.3.1 in [15].

Theorem 2.3 *For any $\tau \geq 4$ the following inequality holds*

$$\mathbf{E} \sup_{u,v \in U} \sup_{s,t \in T} |X(u,t) - X(v,s)| \leq 24(\gamma_{X,U}^\tau(T) + \mathcal{E}_{X,T}^\tau(U)).$$

Proof. We first observe that if $\gamma_{X,U}^\tau(T) < \infty$ then, by the definition, for any $\varepsilon > 0$ there exists an admissible partition sequence $\mathcal{C} = (\mathcal{C}_n)_{n \geq 0}$ of T , which satisfies

$$(1 + \varepsilon)\gamma_{X,U}^\tau(T) \geq \sup_{t \in T} \sum_{n=0}^{\infty} \Delta_{n+\tau,U}(B_n(t)).$$

Let us fix $C \in \mathcal{C}_n$ and let $\pi_n(C)$ be a point in T such that

$$e_{n,\pi_n(C)}^\tau \leq (1 + \varepsilon) \inf\{e_{n,t}^\tau : t \in B\}.$$

Consequently, for any $n \geq 2$ there exists a partition $\mathcal{B}_{C,n-2}$ of the set U into at most N_{n-2} sets B that satisfy

$$\Delta_{n+\tau-2,\pi_{n-2}(C)}(B) \leq 2e_{n-2,\pi_{n-2}(C)}^\tau, \text{ where } \Delta_{n,t}(B) = \sup_{u,v \in B} q_{n,t}(u,v).$$

Using sets $B \times C$ for $B \in \mathcal{B}_{C,n-2}$ and $C \in \mathcal{C}_{n-2}$ we get a partition \mathcal{A}'_{n-2} of $U \times T$ into at most $N_{n-2}^2 \leq N_{n-1}$ sets. Finally intersecting all the constructed sets in $\mathcal{A}'_0, \mathcal{A}'_1, \dots, \mathcal{A}'_{n-2}$ we obtain a nested sequence of partitions $(\mathcal{A}_n)_{n \geq 2}$ such that $|\mathcal{A}_n| \leq N_n$. We complete the sequence by $\mathcal{C}_0 = \mathcal{C}_1 = \{U \times T\}$. In this way $\mathcal{A} = (\mathcal{A}_n)_{n \geq 0}$ is an admissible sequence of partitions for $U \times T$. Let $A_n(u,t)$ be the element of \mathcal{A}_n that contains point (u,t) . Clearly

$$A_n(u,t) \subset B \times C,$$

where $C = C_{n-2}(t)$ and $u \in B \in \mathcal{B}_{C,n-2}$. Therefore for $n \geq 2$,

$$\sup_{s,s' \in C} q_{n+\tau-2,U}(s,s') \leq \Delta_{n+\tau-2,U}(C)$$

and

$$\sup_{v,v' \in B} q_{n+\tau-2,\pi_n(C)}(v,v') \leq 2e_{n-2,\pi_{n-2}(C)}^\tau \leq 2(1 + \varepsilon)e_{n-2,t}^\tau.$$

We turn to the analysis of optimal quantiles \bar{q}_n for the process $X(u,t)$, $u \in \mathcal{A}$, $t \in T$. We show that for any $x, y, z \in T$ and $v, w \in U$

$$\bar{q}_n((v,x), (w,y)) \leq q_{n,U}(x,z) + q_{n,U}(y,z) + q_{n,z}(v,w).$$

This holds due to the triangle inequality

$$\begin{aligned}
& \bar{q}_n((v, x), (w, y)) \\
& \leq \bar{q}_n((v, x), (v, z)) + \bar{q}_n((w, y), (w, z)) + \bar{q}_n((v, z), (w, z)) \\
& \leq q_{n,v}(x, z) + q_{n,w}(y, z) + q_{n,z}(v, w) \\
& \leq q_{n,U}(x, z) + q_{n,U}(y, z) + q_{n,z}(v, w).
\end{aligned}$$

In particular it implies that for any $(v, s) \in B \times C$

$$\begin{aligned}
& \bar{q}_{n+\tau-2}((u, t), (v, s)) \\
& \leq q_{n+\tau-2,U}(t, \pi_{n-2}(C)) + q_{n+\tau-2,U}(s, \pi_{n-2}(C)) + q_{n+\tau-2,\pi_{n-2}(C)}(u, v)
\end{aligned}$$

and hence

$$\Delta_{n+\tau-2}(B \times C) \leq 2\Delta_{n+\tau-2,U}(C_{n-2}(t)) + 2(1 + \varepsilon)e_{n-2,t}^\tau.$$

If $\tau \geq 2$ then also for $n = 0, 1$,

$$\Delta_{n+\tau-2}(U \times T) \leq \Delta_\tau(U \times T) \leq 2\Delta_\tau(T) + 2(1 + \varepsilon)\sup_{t \in T} e_{0,t}^\tau.$$

It implies that for any $\tau \geq 2$

$$\gamma_X^{\tau-2}(U \times T) \leq 4\Delta_\tau(T) + 4(1 + \varepsilon)\sup_{t \in T} e_{0,t}^\tau + 2(1 + \varepsilon)\gamma_{X,U}^\tau(T) + 2(1 + \varepsilon)\mathcal{E}_{X,T}^\tau(U).$$

Therefore for any $\tau \geq 4$ we may apply Theorem 1.3 for distances \bar{q}_n and in this way prove our result with the constant 24. ■

In the next sections we analyse various applications of the result.

3 Gaussian case

As we have mentioned the basic example for the theory are Gaussian canonical processes. Suppose that $T \subset \ell^2(I)$, we recall that

$$G(t) = \sum_{i \in I} t_i g_i, \quad t \in T,$$

where $(g_i)_{i \in I}$ is a sequence of i.i.d. standard normal variables. For the process $G(t)$, $t \in T$ the natural distance is

$$d(s, t) = (\mathbf{E}|G(t) - G(s)|^2)^{\frac{1}{2}} = \|t - s\|_2, \quad s, t \in T.$$

It is easy to see that the optimal quantiles for G satisfy

$$\bar{q}_n(s, t) \sim 2^{\frac{n}{2}} d(s, t), \quad \text{for all } s, t \in T.$$

Consequently denoting $q_n(s, t) = C2^{\frac{n}{2}} d(s, t)$ for large enough C we get an admissible sequence of distances. Moreover,

$$\gamma_G(T) \sim \gamma_2(T, d) = \inf_{\mathcal{A}} \sup_{t \in T} \sum_{n=0}^{\infty} 2^{\frac{n}{2}} \Delta(A_n(t)), \quad (3.1)$$

where the infimum is taken over all admissible $\mathcal{A} = (\mathcal{A}_n)_{n \geq 0}$ sequences of partitions and $\Delta(A) = \sup_{s,t \in A} d(s,t)$. Since d is the canonical distance for $\ell^2(I)$ we usually suppress d from $\gamma_2(T, d)$ and simply write $\gamma_2(T)$. As we have mentioned in the introduction, using (1.3) we get

$$K^{-1}\gamma_2(T) \leq \mathbf{E} \sup_{s,t \in T} |G(t) - G(s)| \leq K\gamma_2(T).$$

Let us also define

$$e_n = \inf\{\varepsilon : N(T, d, \varepsilon) \leq N_n\}, \quad n \geq 0.$$

Obviously $e_n^\tau \leq C2^{\frac{n+\tau}{2}}e_n$ and hence

$$\mathcal{E}_G^\tau(T) \leq C2^{\frac{\tau}{2}} \sum_{n \geq 0} 2^{\frac{n}{2}} e_n.$$

Let $\mathcal{E}(T, d) = \sum_{n=0}^{\infty} 2^{\frac{n}{2}} e_n$, it is the well known fact (Theorem 3.1.1 in [15]) that

$$K^{-1}\mathcal{E}(T, d) \leq \int_0^\infty \sqrt{\log(N(T, d, \varepsilon))} d\varepsilon \leq K\mathcal{E}(T, d). \quad (3.2)$$

Again since d is the canonical distance on $\ell^2(I)$ we will suppress d from $\mathcal{E}(T, d)$ and write $\mathcal{E}(T)$. By (2.3) we infer Dudley's bound

$$\mathbf{E} \sup_{s,t \in T} |G(t) - G(s)| \leq 16C2^{\frac{\tau}{2}} \mathcal{E}(T).$$

Turning our attention to product spaces let us recall that for $U \subset \mathbb{R}^I$ or \mathbb{C}^I , $T \subset \mathbb{R}^I$ or \mathbb{C}^I such that $ut = (u_i t_i)_{i \in I} \in \ell^2(I)$ for all $u \in U$ and $t \in T$ we may define

$$G(u, t) = \sum_{i \in I} u_i t_i g_i, \quad u \in U, t \in T,$$

where $(g_i)_{i \in I}$ is a Gaussian sequence. Note that for all $s, t \in T$, $u, v \in U$

$$\bar{q}_n((u, t), (v, s)) \leq C2^{\frac{n}{2}} \|ut - vs\|_2.$$

For each $u \in U$ and $s, t \in T$ let $d_u(s, t) = \|u(t - s)\|_2$. We may define

$$q_{n,u}(s, t) = C2^{\frac{n}{2}} d_u(s, t), \quad q_{n,U}(s, t) = C2^{\frac{n}{2}} \sup_{u \in U} d_u(s, t).$$

In particular

$$q_{n,U}(s, t) \leq C \sup_{u \in U} \|u\|_\infty d(s, t)$$

and therefore

$$\gamma_{G,U}^\tau(T) \leq C2^{\frac{\tau}{2}} \sup_{u \in U} \|u\|_\infty \gamma_2(T, d). \quad (3.3)$$

On the other hand we define for all $t \in T$ and $u, v \in U$

$$q_{n,t}(u, v) = C2^{\frac{n}{2}} \|t(u - v)\|_2. \quad (3.4)$$

For each $t \in T$ let us denote by d_t the distance on U given by

$$d_t(u, v) = \left(\sum_{i \in I} |t_i|^2 |u_i - v_i|^2 \right)^{\frac{1}{2}}, \quad u, v \in U.$$

Using these distances we may rewrite (3.4) as $q_{n,t} = C2^{\frac{n}{2}}d_t$. Let

$$e_{n,t} = \inf\{\varepsilon : N(U, d_t, \varepsilon) \leq N_n\}, \quad n \geq 0 \quad \text{and} \quad \mathcal{E}(U, d_t) = \sum_{n=0}^{\infty} 2^{\frac{n}{2}} e_{n,t},$$

we obtain by (3.4)

$$e_{n,t}^{\tau} \leq C2^{\frac{n+\tau}{2}} e_{n,t} \quad \text{and} \quad \mathcal{E}_{X,T}^{\tau}(U) \leq C2^{\frac{\tau}{2}} \sup_{t \in T} \mathcal{E}(U, d_t).$$

Using (3.2) we have

$$K^{-1} \mathcal{E}(U, d_t) \leq \int_0^{\infty} \sqrt{\log(N(U, d_t, \varepsilon))} d\varepsilon \leq K \mathcal{E}(U, d_t).$$

We recall also that if $0 \in T$ then by (1.3) $\gamma_2(T) \sim g(T) = \mathbf{E} \sup_{t \in T} |G(t)|$. We may state the following corollary of Theorem 2.3 which extends slightly Theorem 3.3.1 in [15].

Corollary 3.1 *For any $\tau \geq 4$*

$$\begin{aligned} & \mathbf{E} \sup_{u,v \in U} \sup_{s,t \in T} |G(u,t) - G(v,s)| \\ & \leq 32(\gamma_{G,U}^{\tau}(T) + \mathcal{E}_{G,T}^{\tau}(U)) \leq 32C2^{\frac{\tau}{2}} \left(\sup_{u \in U} \|u\|_{\infty} \gamma_2(T) + \sup_{t \in T} \mathcal{E}(U, d_t) \right). \end{aligned}$$

Moreover

$$\mathcal{E}(U, d_t) \sim \int_0^{\infty} \sqrt{\log N(U, d_t, \varepsilon)} d\varepsilon \quad \text{and} \quad \gamma_2(T) \sim g(T) \quad \text{if } 0 \in T.$$

It is a tempting idea to replace $\sup_{u \in U} \|u\|_{\infty} \gamma_2(T)$ and $\sup_{t \in T} \mathcal{E}(U, d_t)$ by respectively $\sup_{u \in U} \gamma_2(T, d_u)$ and $\sup_{t \in T} \gamma_2(U, d_t)$. We show that this approach cannot work, to this aim lets us consider toy example where T and U are usual ellipsoids, i.e.

$$U = \left\{ u \in \mathbb{R}^I : \sum_{i \in I} \frac{|u_i|^2}{|x_i|^2} \leq 1 \right\}. \quad (3.5)$$

and

$$T = \left\{ t \in \mathbb{R}^I : \sum_{i \in I} \frac{|t_i|^2}{|y_i|^2} \leq 1 \right\}.$$

Obviously

$$\mathbf{E} \sup_{u \in U, t \in T} |G(u, t)| = \mathbf{E} \max_{i \in I} |x_i y_i g_i|$$

on the other hand

$$\sup_{u \in U} \mathbf{E} \sup_{t \in T} |G(u, t)| \sim \sup_{u \in U} \|uy\|_2 = \max_{i \in I} |x_i y_i| = \|xy\|_{\infty}.$$

Similarly, $\sup_{t \in T} \mathbf{E} \sup_{u \in U} |G(u, t)| \sim \|xy\|_{\infty}$. However $\|xy\|_{\infty} \leq 1$ does not guarantee that $\mathbf{E} \max_{i \in I} |x_i y_i g_i|$ is finite, for example if $x_i = y_i = 1$. On the other hand Corollary 3.1 implies the following result for the ellipsoid U .

Remark 3.2 Suppose that \mathcal{A} is given by (3.5). Then for any set $0 \in T \subset \ell^2(I)$,

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |G(u, t)| \leq K(\|x\|_\infty g(T) + \Delta(T)\|x\|_2).$$

Proof. For the sake of simplicity we assume that $I = \mathbb{N}$. We use Corollary 3.1 together with the following observation. For each $t \in T$ points $(u_i t_i)_{i \in \mathbb{N}}$, $u \in U$ forms an ellipsoid $U_t = \{a \in \mathbb{R}^\mathbb{N} : \sum_{i \in \mathbb{N}} \frac{|u_i|^2}{|t_i|^2 |x_i|^2} \leq 1\}$ and therefore by Proposition 2.5.2 in [15]

$$\int_0^\infty \sqrt{\log N(U, d_t, \varepsilon)} d\varepsilon \leq L \sum_{n=0}^\infty 2^{\frac{n}{2}} |x_{2^n} t_{2^n}|,$$

where L is a universal constant, assuming that we have rearranged $(x_i t_i)_{i \in \mathbb{N}}$ in such a way that the sequence $|x_i t_i|$, $i \in \mathbb{N}$ is non-increasing. It suffices to note that by the Schwartz inequality

$$\begin{aligned} \sum_{n=0}^\infty 2^{\frac{n}{2}} (|x_{2^n} t_{2^n}|^2) &\leq 2|x_1 t_1| + 2 \sum_{n=1}^\infty \left(\sum_{i=2^{n-1}+1}^{2^n} |x_i t_i|^2 \right)^{\frac{1}{2}} \\ &\leq 2|x_1 t_1| + 2 \sum_{n=1}^\infty \max_{2^{n-1} < i \leq 2^n} |x_i| \left(\sum_{i=2^{n-1}+1}^{2^n} |t_i|^2 \right)^{\frac{1}{2}} \leq 2\|x\|_2 \|t\|_2. \end{aligned}$$

Consequently,

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |G_{t,u}| \leq K(\|x\|_\infty \gamma_2(T) + \Delta(T)\|x\|_2).$$

It remains to apply (1.3), i.e. $g(T) \sim \gamma_2(T)$. ■

Note that $\|x\|_\infty = \sup_{u \in U} \|u\|_\infty$ and $\|x\|_2 \sim \gamma_2(U, d)$. It is clear the result can be slightly improved if $\mathcal{E}(U, d) < \infty$. Indeed similarly to (3.3) one can show that

$$\mathcal{E}_{G,U}^\tau(T) \leq C 2^{\frac{\tau}{2}} \sup_{t \in T} \|t\|_\infty \mathcal{E}(U, d)$$

and hence by Corollary 3.1 for any set U such that $\mathcal{E}(U, d) < \infty$

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |G(u, t)| \leq K \left(\|x\|_\infty g(T) + \sup_{t \in T} \|t\|_\infty \right).$$

This is a sort of general rule that we expect certain regularity of \mathcal{A} , whereas set T is usually supposed to be unknown.

We turn to show that Corollary 3.1 is sufficiently strong to answer both questions about VC classes and Fourier series, posed in the introduction, in the Gaussian type formulation. Namely we prove that these problems are related to certain properties of entropy functionals $\int_0^\infty \sqrt{\log(N(\mathcal{A}, d_t, \varepsilon))} d\varepsilon$, $t \in T$.

We start from the result on VC classes of finite dimension. In this case \mathcal{A} consists of $a \in \mathbb{R}^I$, of the form $a = 1_A$ for some $A \subset I$ and T is any subset of

$\ell^2(I)$ that contains 0. By Theorem 14.12 in [11] we have that for any $t \in \ell_2(I)$ such that $\|t\|_2 = 1$ and given VC class \mathcal{A} of dimension d

$$\log N(\mathcal{A}, d_t, \varepsilon) \leq Ld \left(1 + \log \frac{1}{\varepsilon}\right), \quad 0 < \varepsilon < 1.$$

Consequently

$$\sup_{t \in T} \int_0^\infty \sqrt{\log N(\mathcal{A}, d_t, \varepsilon)} d\varepsilon \leq \sqrt{Ld} \Delta(T) \int_0^1 (1 + \log \frac{1}{\varepsilon})^{\frac{1}{2}} d\varepsilon \leq M \sqrt{d} \Delta(T)$$

and hence $\mathcal{E}(\mathcal{A}, d_t) \leq M \sqrt{d} \Delta(T)$ for any $t \in T$. Since clearly $\Delta(T) \leq \gamma_2(T)$ we may use Corollary 3.1 for $\tau = 4$ and deduce

$$\mathbf{E} \sup_{a \in \mathcal{A}} \sup_{t \in T} |G(a, t)| \leq 64C(\gamma_2(T) + M \sqrt{d} \Delta(T)) \leq L \sqrt{d} \gamma_2(T),$$

which by (1.3) implies (1.5) - the result we refer to.

Next step is to show that the Gaussian version of the problem on Fourier series is also related to Corollary 3.1. In this case U consists of $u \in \mathbb{C}^I$ of the form $u_i = \chi_i(h)$ for $h \in G$, where $\chi_i, i \in I$ are characters on the compact Abelian group G and T is any subset of $\ell^2(I)$ that contains 0. Recall that the crucial observation for the study is that distances $d_t, t \in T$ defined in (1.12) are translationally invariant on the group G , i.e.

$$d_t(f \cdot h, g \cdot h) = d_t(f, g) = \left(\sum_{i \in I} |t_i|^2 |\chi_i(f) - \chi_i(g)|^2 \right)^{\frac{1}{2}}, \quad \text{for any } f, g, h \in G.$$

Therefore by the deep result of Fernique (Theorem 3.1.1 in [15]) we have

$$K^{-1} \mathcal{E}(\mathcal{A}, d_t) \leq \mathbf{E} \sup_{a \in \mathcal{A}} |G(a, t)| \leq K \mathcal{E}(\mathcal{A}, d_t).$$

Consequently by Corollary 3.1 and (1.3) we can deduce (1.7) which is the Gaussian version of Theorem 1.3.

The first new consequence of Corollary 3.1 concerns shattering dimension of U . Suppose that U is the class of real functions bounded by 1, i.e. $U \subset [-1, 1]^I$. We say that a subset B of I is ε -shattered if there exists a level function v on B such that given any subset A of B one can find a function $u \in U$ with $u_i \leq v_i$ if $i \in A$ and $u_i \geq v_i + \varepsilon$ if $i \in B \setminus A$. The shattering dimension of U denoted by $\text{vc}(U, \varepsilon)$ is the maximal cardinality of a set ε -shattered by U . The deep result of Mendelson and Vershynin [14] is that for any $t \in \ell_2(I)$ such that $\|t\|_2 = 1$ we have

$$\log N(U, d_t, \varepsilon) \leq L \text{vc}(U, c\varepsilon) \log \left(\frac{2}{\varepsilon} \right), \quad 0 < \varepsilon < 1, \quad (3.6)$$

where L and c are positive absolute constants. This leads to the following corollary.

Corollary 3.3 *Suppose that $0 \in T \subset \ell^2(I)$ and $\sup_{u \in U} \|u\|_\infty$. The following inequality holds*

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |G(u, t)| \leq K \left(g(T) + \Delta(T) \int_0^1 \sqrt{\text{vc}(U, \varepsilon) \log(2/\varepsilon)} d\varepsilon \right).$$

Proof. It suffices to apply (3.6) then Corollary 3.1 and finally (1.3). ■

It is worth mentioning how to understand the shattering dimension of convex bodies. Suppose that $U \subset [-1, 1]^d$ is a convex and symmetric body then $\text{vc}(U, \varepsilon)$ is the maximal cardinality of a subset J of $\{1, 2, \dots, d\}$ such that $P_J(U) \supset [-\frac{\varepsilon}{2}, \frac{\varepsilon}{2}]^J$, where P_J is the orthogonal projection from \mathbb{R}^d on \mathbb{R}^J . For example suppose that U is a unit ball in \mathbb{R}^d then $\text{vc}(U, \varepsilon) = k - 1$ for any $\varepsilon \in [\frac{2}{\sqrt{k}}, \frac{2}{\sqrt{k-1}})$ and $k = 1, 2, \dots, d$ moreover $\text{vc}(U, \varepsilon) = d$ for $\varepsilon < \frac{2}{\sqrt{d}}$. Consequently

$$\int_0^1 \sqrt{\text{vc}(U, \varepsilon) \log(2/\varepsilon)} d\varepsilon \leq K \sqrt{d \log d}.$$

Note that for $t \in \mathbb{R}^d$, $t_i = 1/\sqrt{d}$ we have that $\int_0^\infty \sqrt{\log N(U, d_t, \varepsilon)} d\varepsilon$ is comparable to \sqrt{d} and hence the above estimate is not far from this answer.

4 Bernoulli case

Our next aim is to obtain a version of Corollary 3.1 in the setting of Bernoulli processes. We recall that by Bernoulli processes we mean

$$X(t) = \sum_{i \in I} t_i \varepsilon_i, \text{ for } t \in T \subset \ell^2(I),$$

where $(\varepsilon_i)_{i \in I}$ is a Bernoulli sequence. Moreover we denoted

$$b(T) = \mathbf{E} \sup_{t \in T} |X(t)|$$

and if $0 \in T$ then by Theorem 1.1 we have a geometrical characterization of $b(T)$. We turn to the analysis of Bernoulli processes on product spaces, namely we consider

$$X(u, t) = \sum_{i \in I} u_i t_i \varepsilon_i, \quad t \in T, u \in U,$$

where $T \subset \mathbb{R}^I$ or \mathbb{C}^I , $U \subset \mathbb{R}^I$ or \mathbb{C}^I and $(\varepsilon_i)_{i \in I}$ is a Bernoulli sequence. Our approach is to use Theorem 1.1 in order to extend Corollary 3.1 to the case of random signs.

Theorem 4.1 *Suppose that $0 \in T$. Then there exists $\pi : T \rightarrow \ell^2$ such that $\|\pi(t)\|_1 \leq Lb(T)$ for all $t \in T$, $\pi(0) = 0$ and*

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| \leq K \left(\sup_{u \in U} \|u\|_\infty b(T) + \sup_{t \in T} \mathcal{E}(U, d_{t-\pi(t)}) \right),$$

where

$$\mathcal{E}(U, d_{t-\pi(t)}) \sim \int_0^\infty \sqrt{\log N(U, d_{t-\pi(t)}, \varepsilon)} d\varepsilon.$$

Proof. Obviously we may assume that $b(T) = \mathbf{E} \sup_{t \in T} |\sum_{i \in I} t_i \varepsilon_i| < \infty$. Therefore by Theorem 1.1 there exists a decomposition $T \subset T_1 + T_2$, $T_1, T_2 \subset \ell_2(I)$ which satisfies $0 \in T_2$ and

$$\max \left\{ \sup_{t \in T_1} \|t\|_1, g(T_2) \right\} \leq Lb(T), \quad (4.1)$$

where L is a universal constant. Hence combining Corollary 3.1 with (4.1)

$$\begin{aligned}
& \mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| \\
& \leq C \left(\mathbf{E} \sup_{u \in U} \sup_{t \in T_1} |X(u, t)| + \mathbf{E} \sup_{u \in U} \sup_{t \in T_2} |X(u, t)| \right) \\
& \leq C \left(\sup_{u \in U} \|u\|_\infty \sup_{t \in T_1} \|t\|_1 + \sqrt{\frac{\pi}{2}} \mathbf{E} \sup_{u \in U} \sup_{t \in T_2} |G(u, t)| \right) \\
& \leq CL \left(\sup_{u \in U} \|u\|_\infty b(T) + \sup_{t \in T_2} \int_0^\infty \sqrt{\log N(U, d_t, \varepsilon)} d\varepsilon \right).
\end{aligned}$$

The decomposition of T into $T_1 + T_2$ can be defined in a way that $T_1 = \{\pi(t) : t \in T\}$ and $T_2 = \{t - \pi(t) : t \in T\}$ where $\pi : T \rightarrow \ell_2$ is such that $\pi(0) = 0$, $\|\pi(t)\|_1 \leq Lb(T)$ and $\gamma_2(T_2) \leq Lb(T)$. It completes the proof. \blacksquare

Our remarks on the entropy function from the previous chapter shows that Theorem 4.1 solves both our questions from the introduction. In fact we can easily extend our result for the functional shattering dimension.

Corollary 4.2 *Suppose that $\sup_{u \in U} \|u\|_\infty \leq 1$ and U is of ε -shattering dimension $\text{vc}(\mathcal{A}, \varepsilon)$ for $0 < \varepsilon < 1$, then*

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| \leq Kb(T) \int_0^1 \sqrt{\text{vc}(U, \varepsilon) \log(2/\varepsilon)} d\varepsilon.$$

Proof. Obviously by Theorem 4.1,

$$\begin{aligned}
& \mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| \\
& \leq K \left(\sup_{u \in U} \|u\|_\infty b(T) + \sup_{t \in T} \|t - \pi(t)\|_2 \int_0^1 \sqrt{\text{vc}(U, \varepsilon) \log(2/\varepsilon)} d\varepsilon \right),
\end{aligned}$$

where $\|\pi(t)\|_1 \leq Lb(T)$. Since

$$\sup_{t \in T} \|t - \pi(t)\|_2 \leq Lb(T) \quad \text{and} \quad \sup_{u \in U} \|u\|_\infty \leq 1$$

it implies that

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| \leq Kb(T) \int_0^1 \sqrt{\text{vc}(U, \varepsilon) \log(2/\varepsilon)} d\varepsilon. \quad \blacksquare$$

Corollary 4.3 *Suppose that $\sup_{u \in U} \|u\|_\infty \leq 1$ and $v_i \in F$, $i \in I$, then*

$$\mathbf{E} \sup_{u \in U} \left\| \sum_{i \in I} u_i v_i \varepsilon_i \right\| \leq K \mathbf{E} \left\| \sum_{i \in I} v_i \varepsilon_i \right\| \int_0^1 \sqrt{\text{vc}(U, \varepsilon) \log(2/\varepsilon)} d\varepsilon.$$

Note that in the same way as for the maximal inequality we may ask what is the right characterization of $U \subset \mathbb{R}^I$, $\sup_{u \in U} \|u\|_\infty \leq 1$ for which the inequality

$$\mathbf{E} \sup_{u \in U} \left\| \sum_{i \in I} u_i X_i \right\| \leq K \left\| \sum_{i \in I} X_i \right\|, \quad (4.2)$$

holds for any sequence of independent symmetric r.v.'s X_i , $i \in I$ which take values in a separable Banach space F . With the same proof as the first part of Theorem 1.2 one can show that U should satisfy the condition

$$\sup_{0 < \varepsilon < 1} \varepsilon \sqrt{\text{vc}(U, \varepsilon)} < \infty.$$

On the other hand Corollary 4.3 implies that

$$\int_0^1 \sqrt{\text{vc}(U, \varepsilon) \log(2/\varepsilon)} d\varepsilon < \infty$$

is a sufficient condition for the inequality (4.2). The exact answer for this question seems to stay at the moment beyond our reach.

We revisit our toy example where U is the ellipsoid

$$U = \left\{ u \in \mathbb{R}^I : \sum_{i \in I} \frac{|u_i|^2}{|x_i|^2} \leq 1 \right\},$$

where $|x_i| > 0$ are given numbers. One can show the following result.

Remark 4.4 Suppose that U is the usual ellipsoid. Then for any set $0 \in T \subset \ell^2(I)$

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| \leq K \|x\|_2 b(T).$$

Proof. We may argue in a similar way as in Remark 3.2 using this time Theorem 4.1 and obtain

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| \leq K \left(\|x\|_\infty b(T) + \sup_{t \in T} \|t - \pi(t)\|_2 \|x\|_2 \right).$$

Since $\Delta(T) \leq Lb(T)$ and $\|\pi(t)\|_1 \leq Lb(T)$ we get $\sup_{t \in T} \|t - \pi(t)\|_2 \leq Lb(T)$ and therefore

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| \leq K \|x\|_2 b(T).$$

■

On the other hand by the Schwartz inequality

$$\mathbf{E} \sup_{u \in U} \sup_{t \in T} |X(u, t)| = \sup_{t \in T} \mathbf{E} \sup_{u \in U} |X(u, t)| = \sup_{t \in T} \|tx\|_2$$

Consequently in this case the expectation of the supremum of the Bernoulli process over the product of index sets can be explained by the analysis of one of its marginal processes.

It is worth mentioning that whenever there holds a comparability of moments like (4.2) one can deduce also the comparability of tails.

Remark 4.5 Suppose that $\sup_{u \in U} \|u\|_\infty \leq 1$. If for any X_i , $i \in I$ symmetric independent random variables which take values in a separable Banach space $(F, \|\cdot\|)$

$$\mathbf{E} \sup_{u \in U} \left\| \sum_{i \in I} u_i X_i \right\| \leq L \mathbf{E} \left\| \sum_{i \in I} X_i \right\|, \quad (4.3)$$

then there exists an absolute constant K such that

$$\mathbf{P} \left(\sup_{u \in U} \left\| \sum_{i \in I} u_i X_i \right\| \geq Kt \right) \leq K \mathbf{P} \left(\left\| \sum_{i \in I} X_i \right\| \geq t \right).$$

Proof. It suffices to prove the inequality for $X_i = v_i \varepsilon_i$, where $v_i \in F$ and ε_i are independent Bernoulli variables. The general result follows then from the Fubini theorem. By the result of Dilworth and Montgomery-Smith [2] for all $p \geq 1$

$$\begin{aligned} & \left\| \sup_{u \in U} \left\| \sum_{i \in I} u_i v_i \varepsilon_i \right\| \right\|_p \\ & \leq C \left(\left\| \sup_{u \in U} \left\| \sum_{i \in I} u_i v_i \varepsilon_i \right\| \right\|_1 + \sup_{u \in U} \sup_{\|x^*\| \leq 1} \left\| \sum_{i \in I} u_i x_i^*(v_i) \varepsilon_i \right\|_p \right). \end{aligned}$$

Therefore by the Markov inequality and the assumption $\sup_{u \in U} \|u\|_\infty \leq 1$ we obtain for all $p \geq 1$

$$\begin{aligned} & \mathbf{P} \left(\sup_{u \in U} \left\| \sum_{i \in I} u_i v_i \varepsilon_i \right\| \geq C \left(\mathbf{E} \left\| \sum_{i \in I} v_i \varepsilon_i \right\| + \sup_{\|x^*\| \leq 1} \left\| \sum_{i \in I} x^*(v_i) \varepsilon_i \right\|_p \right) \right) \\ & \leq e^{-p}. \end{aligned} \tag{4.4}$$

On the other hand it is known (e.g. [8]) that for any functional $x^* \in F^*$

$$\mathbf{P} \left(\left| \sum_{i \in I} x^*(v_i) \varepsilon_i \right| \geq M^{-1} \left\| \sum_{i \in I} x^* v_i \varepsilon_i \right\|_p \right) \geq \min \{c, e^{-p}\}, \tag{4.5}$$

where $M \geq 1$. Hence

$$\begin{aligned} & \mathbf{P} \left(\sup_{u \in U} \left\| \sum_{i \in I} u_i v_i \varepsilon_i \right\| \geq C \left(\mathbf{E} \left\| \sum_{i \in I} v_i \varepsilon_i \right\| + Mt \right) \right) \\ & \leq \sup_{\|x^*\| \leq 1} c^{-1} \mathbf{P} \left(\left| \sum_{i \in I} x^*(v_i) \varepsilon_i \right| > t \right) \leq c^{-1} \mathbf{P} \left(\left\| \sum_{i \in I} v_i \varepsilon_i \right\| > t \right). \end{aligned}$$

We end the proof considering two cases. If $t \geq M^{-1} \mathbf{E} \left\| \sum_{i \in I} v_i \varepsilon_i \right\|$ then

$$\mathbf{P} \left(\sup_{u \in U} \left\| \sum_{i \in I} u_i v_i \varepsilon_i \right\| \geq 2C Mt \right) \leq c^{-1} \mathbf{P} \left(\left\| \sum_{i \in I} v_i \varepsilon_i \right\| > t \right).$$

If $t \leq M^{-1} \mathbf{E} \left\| \sum_{i \in I} v_i \varepsilon_i \right\|$ then by Paley-Zygmund inequality and the Kahane inequality with the optimal constant [6]

$$\mathbf{P} \left(\left\| \sum_{i \in I} v_i \varepsilon_i \right\| > t \right) \geq M^{-2} \frac{(\mathbf{E} \left\| \sum_{i \in I} \varepsilon_i v_i \right\|)^2}{\mathbf{E} \left\| \sum_{i \in I} \varepsilon_i v_i \right\|^2} \geq \frac{1}{2} M^{-2}.$$

It shows that

$$\mathbf{P} \left(\sup_{u \in U} \left\| \sum_{i \in I} u_i v_i \varepsilon_i \right\| \geq Kt \right) \leq K \mathbf{P} \left(\left\| \sum_{i \in I} v_i \varepsilon_i \right\| > t \right),$$

which completes the proof. ■

5 Applications

First application we show concerns processes on $\mathcal{A} = [0, 1]$ but with values in \mathbb{R}^I . In order to discuss whether or not $X(a)$, $a \in \mathcal{A}$ has its paths in $\ell^2(I)$ space we have to verify the condition $\|X(a)\|_2 < \infty$ a.s. for all $a \in \mathcal{A}$. This is the same question as

$$\sup_{\|x^*\| \leq 1} \sup_{a \in \mathcal{A}} \langle x^*, X(a) \rangle < \infty.$$

Hence it suffices to check the condition $\mathbf{E} \sup_{a \in \mathcal{A}} \sup_{t \in T} |X(a, t)| < \infty$, where $T = \{t \in \ell^2(I) : \|t\|_2 \leq 1\}$ and

$$X(a, t) = \langle t, X(a) \rangle, \quad t \in T, a \in \mathcal{A}.$$

Note that in this way we match each random vector $X(a)$ with the process $X(a, t)$, $t \in T$. By Theorem 2.3 (note that $X(0, s) = 0$)

$$\mathbf{E} \sup_{a \in \mathcal{A}} \sup_{t \in T} |X(a, t)| \leq 32(\gamma_{X, \mathcal{A}}^\tau(T) + \mathcal{E}_{X, T}^\tau(\mathcal{A})).$$

In the described setting we usually expect that

$$q_{n, a}(s, t) \leq q_{n, 1}(s, t), \quad a \in \mathcal{A} \tag{5.1}$$

which means that on average the increments increases with time. Condition (5.1) yields $\gamma_{X, \mathcal{A}}^\tau(T) \leq \gamma_{X(1)}^\tau(T)$. Under certain assumptions we may also expect that $\gamma_{X(1)}^\tau(T)$ is equivalent to $\mathbf{E}\|X(1)\|_2$. For example this is the case when $X(1)$ is a centred Gaussian vector and also if $X(1)$ consists of entries $X(1)_i$, $i \in I$ that are independent random variables that satisfy some technical assumptions as stated in [10]. Moreover if there exists increasing family of functions $\eta_n : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ which are continuous, increasing, $\eta_n(0) = 0$ such that for all $t \in T$ and $a, b \in \mathcal{A}$

$$q_{n, t}(a, b) \leq \eta_n(|a - b|) \quad \text{and} \quad \sum_{n=0}^{\infty} \eta_{n+\tau}^{-1}(N_n) < \infty \tag{5.2}$$

then

$$\mathcal{E}_{X, T}^\tau(\mathcal{A}) \leq \sum_{n=0}^{\infty} \eta_{n+\tau}^{-1}(N_n) < \infty.$$

In this way we obtain the following remark which is a generalization results from [3, 4].

Remark 5.1 Suppose that $X(a, t)$, $a \in \mathcal{A}$, $t \in T$ satisfies (5.1), (5.2) and $\mathbf{E}\|X(1)\|_2$ is comparable with $\gamma_{X(1)}(T)$ then

$$\mathbf{E} \sup_{a \in \mathcal{A}} \sup_{t \in T} |X(a, t)| \leq K(\mathbf{E}\|X(1)\|_2 + 1).$$

The second application we would like to discuss concerns empirical processes. Let $(\mathcal{X}, \mathcal{B})$ be a measurable space, \mathcal{F} be a countable family of measurable functions $f : \mathcal{X} \rightarrow \mathbb{R}$ such that $0 \in \mathcal{F}$ and X_1, X_2, \dots, X_N be family of independent random variables that satisfy

$$\mathbf{P} \left(|(f - g)(X_i)| > d_1(f, g)t + d_2(f, g)t^{\frac{1}{2}} \right) \leq 2 \exp(-t), \text{ for all } t > 0. \quad (5.3)$$

We will analyse the case when $d_2(f, g) \geq d_1(f, g)$. By Exercise 9.3.5 in [15] for any centred independent random variables Y_1, Y_2, \dots, Y_N which satisfy

$$\mathbf{P} \left(|Y_i| \geq At + Bt^{\frac{1}{2}} \right) \leq 2 \exp(-t) \text{ for all } i = 1, 2, \dots, N,$$

where $A \leq B$ and for any numbers u_1, u_2, \dots, u_N we have

$$\mathbf{P} \left(\left| \sum_{i=1}^N u_i Y_i \right| \geq L \left(A\|u\|_{\infty} t + B\|u\|_2 t^{\frac{1}{2}} \right) \right) \leq 2 \exp(-t) \text{ for all } t > 1. \quad (5.4)$$

Observe that if we define

$$X(u, f) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \varepsilon_i u_i f(X_i), \quad u \in U, f \in \mathcal{F},$$

where U is a unit ball in \mathbb{R}^N and $(\varepsilon_i)_{i=1}^N$ is a Bernoulli sequence independent of X_i , $i = 1, 2, \dots, N$ then

$$\mathbf{E} \sup_{u \in U} \sup_{f \in \mathcal{F}} |X(u, f)| = \mathbf{E} \sup_{f \in \mathcal{F}} \left(\frac{1}{N} \sum_{i=1}^N |f(X_i)|^2 \right)^{\frac{1}{2}}.$$

Clearly by (5.4) used to $Y_i = \varepsilon_i(f - g)(X_i)$ we get for all $t > 1$

$$\begin{aligned} \mathbf{P} \left(|X(u, f) - X(u, g)| \geq \frac{L}{\sqrt{N}} \left(d_1(f, g)\|u\|_{\infty} t + d_2(f, g)\|u\|_2 t^{\frac{1}{2}} \right) \right) \\ \leq 2 \exp(-t). \end{aligned} \quad (5.5)$$

Thus in particular we can use

$$q_{n,u}(f, g) = L(d_1(f, g)\|u\|_{\infty} 2^n + d_2(f, g)\|u\|_2 2^{\frac{n}{2}}).$$

Then

$$q_{n,U}(f, g) = \frac{L}{\sqrt{N}} (d_1(f, g) 2^n + d_2(f, g) 2^{\frac{n}{2}}). \quad (5.6)$$

Let

$$\gamma_i(\mathcal{F}, d_i) = \inf_{\mathcal{A}} \sup_{f \in \mathcal{F}} \sum_{n=0}^{\infty} 2^{\frac{n}{i}} \Delta_i(A_n(f)),$$

where the infimum is taken over all admissible partitions $\mathcal{A} = (\mathcal{A}_n)_{n \geq 0}$ and $\Delta_i(A) = \sup_{f, g \in A} d_i(f, g)$. It is easy to construct an admissible partitions which works for both $\gamma_1(\mathcal{F}, d_1)$ and $\gamma_2(\mathcal{F}, d_2)$. Namely we consider partitions $\mathcal{A}^1 = (\mathcal{A}_n^1)_{n \geq 0}$ and $\mathcal{A}^2 = (\mathcal{A}_n^2)_{n \geq 0}$ such that

$$(1 + \varepsilon)\gamma_2(\mathcal{F}, d_i) \geq \sup_{f \in \mathcal{F}} \sum_{n=0}^{\infty} 2^{\frac{n}{i}} \Delta_i(\mathcal{A}_n^i(f)), \quad i = 1, 2,$$

for some arbitrary small $\varepsilon > 0$ and then define $\mathcal{A} = (\mathcal{A}_n)_{n \geq 0}$ by $\mathcal{A}_n = \mathcal{A}_{n-1}^1 \cap \mathcal{A}_{n-1}^2$ for $n \geq 1$ and $\mathcal{A}_0 = \{\mathcal{F}\}$. Obviously \mathcal{A} is admissible, moreover,

$$\sup_{f \in \mathcal{F}} \sum_{n=0}^{\infty} 2^{\frac{n}{i}} \Delta_i(\mathcal{A}_n(f)) \leq (1 + \varepsilon) 2^{\frac{1}{i}} \gamma_i(\mathcal{F}, d_i), \quad i = 1, 2. \quad (5.7)$$

Using the partition \mathcal{A} we derive from (5.6)

$$\gamma_{X, \mathcal{A}}^\tau(\mathcal{F}) \leq \frac{2^{\tau+1}(1 + \varepsilon)}{\sqrt{N}} (\gamma_1(\mathcal{F}, d_1) + \gamma_2(\mathcal{F}, d_2)).$$

On the other hand using that $X(u, f) - X(v, f) = X(u - v, f) - X(0, f)$ similarly to (5.5) we get for all $t > 1$

$$\begin{aligned} \mathbf{P} \left(|X(u, f) - X(v, f)| \geq \frac{L}{\sqrt{N}} \left(d_1(f, 0) \|u - v\|_\infty t + d_2(f, 0) \|u - v\|_2 t^{\frac{1}{2}} \right) \right) \\ \leq 2 \exp(-t). \end{aligned}$$

Hence we may define

$$q_{n, f}(u, v) = \frac{L}{\sqrt{N}} \left(d_1(f, 0) \|u - v\|_\infty 2^n + d_2(f, 0) \|u - v\|_2 2^{\frac{n}{2}} \right). \quad (5.8)$$

We aim to compute entropy numbers $e_{n, f}$, $n \geq 0$. Let $n_0 \geq 1$ be such that $2^{n_0} \geq N \geq 2^{n_0-1}$. Our first claim is that for any $n \geq n_0$ and suitably chosen $\sigma \geq 1$ it is possible to cover $B_N(0, 1)$, a unit ball in \mathbb{R}^N , by at most $N_{n+\sigma}$ cubes of ℓ^∞ diameter at most $N^{-\frac{1}{2}}(N_n)^{-\frac{1}{\sigma}}$. Indeed we can apply the volume type argument. It is possible to cover $B_N(0, 1)$ with M disjoint cubes of ℓ^∞ diameter t with disjoint interiors in $B_N(0, 1 + t\sqrt{N})$. Since $|B_N(0, 1)| \sim (C/\sqrt{N})^N$ we get

$$M \leq \frac{|B_N(0, 1 + t\sqrt{N})|}{t^N} \leq \left(\frac{C}{\sqrt{N}} \left(\frac{1 + t\sqrt{N}}{t} \right) \right)^N \leq C^N \left(1 + \frac{1}{t\sqrt{N}} \right)^N.$$

We may choose $t = N^{-\frac{1}{2}} N_n^{-\frac{1}{\sigma}}$, then

$$C^N \left(1 + \frac{1}{t\sqrt{N}} \right)^N = C^N \left(1 + N_n^{\frac{1}{\sigma}} \right)^N \leq (N_{n+\sigma})^{\frac{1}{\sigma}},$$

where the last inequality uses the assumption that $n \geq n_0$. In this way we covered $B_N(0, 1)$ by at most $N_{n+\sigma}$ sets of ℓ_∞ diameter at most $N^{-\frac{1}{2}}(N_n)^{-\frac{1}{\sigma}}$ and ℓ_2 diameter at most $N_n^{-\frac{1}{\sigma}}$. By (5.8) we infer the following entropy bound

$$e_{n+\sigma, f}^\tau \leq L 2^\tau \left(d_1(f, 0) N^{-1} (N_n)^{-\frac{1}{\sigma}} + d_2(f, 0) N^{-\frac{1}{2}} (N_n)^{-\frac{1}{\sigma}} \right).$$

We recall that constant L is absolute but may change its value from line to line up to a numerical factor. This implies the bound

$$\begin{aligned} & \sum_{n=n_0}^{\infty} e_{n+\sigma, f}^{\tau} \\ & \leq L2^{\tau} \left(d_1(f, 0)N^{-1}(N_n)^{-\frac{1}{N}} + d_2(f, 0)N^{-\frac{1}{2}}(N_n)^{-\frac{1}{N}} \right) \\ & \leq L2^{\tau} (d_1(f, 0) + d_2(f, 0)). \end{aligned}$$

The second step is to consider $n \leq n_0/2 + \sigma$. In this case we can simply use the trivial covering of $B_N(0, 1)$ by a single set which obviously has ℓ^{∞} and ℓ^2 diameter equal 2 and hence

$$e_{n, f}^{\tau} = \frac{L2^{\tau}}{\sqrt{N}} (d_1(f, 0)2^n + d_2(f, 0)2^{\frac{n}{2}})$$

and

$$\sum_{n=0}^{n_0/2+\sigma} e_{n, f}^{\tau} \leq L2^{\tau} (d_1(f, 0) + d_2(f, 0)).$$

The most difficult case is when $n_0/2 + \sigma \leq n \leq n_0 + \sigma$. In this setting we will cover $B_N(0, 1)$ with cubes of ℓ^{∞} diameter $2t$, where on $t = \frac{1}{\sqrt{m_n}}$ and $m_n \leq N$. We will not control ℓ^2 diameter, we simply use that it is always bounded by 2. Note that if $x \in B_N(0, 1)$ then only on at most m_n coordinates such that $|x_i| > t$. Therefore we can cover $B_N(0, 1)$ with cubes in \mathbb{R}^N of ℓ^{∞} diameter $2t$ if for each subset $J \subset \{1, 2, \dots, N\}$ such that $|J| = m_n$ we cover $B_J(0, 1) \subset \mathbb{R}^J$ with cubes in \mathbb{R}^J of ℓ^{∞} diameter $2t$. In this way we cover all the cases where all coordinates but those in J stay in the cube $[-t, t]^{J^c}$. By our volume argument one needs at most

$$M_J \leq C^J \left(1 + \frac{2t}{\sqrt{|J|}} \right)^{|J|} = C^{m_n} \left(1 + \frac{1}{2t\sqrt{m_n}} \right)^{m_n}$$

$2t$ -cubes in \mathbb{R}^J to cover $B_J(0, 1)$. Our choice of t guarantees that $M_J \leq (3c/2)^{m_n}$. Therefore one needs $\binom{N}{m_n} (3c/2)^{m_n}$ cubes of ℓ^{∞} diameter $2t$ to cover $B_N(0, 1)$. We require that

$$\binom{N}{m_n} \left(\frac{3c}{2} \right)^{m_n} \leq N_{n+\sigma}.$$

It remains to find m_n that satisfies the above inequality. First observe that

$$\binom{N}{m_n} \leq \left(\frac{eN}{m_n} \right)^{m_n} = \exp(m_n \log(eN/m_n)).$$

Following Talagrand we define m_n as the smallest integer such that

$$2^{n-\sigma} \leq m_n \log(eN/m_n).$$

Clearly if $n \geq n_0/2 + \sigma$ then $m_n > 1$ and thus by Lemma 9.3.12 in [15] we deduce that $m_n \log(eN/m_n) \leq 2^{n-\sigma+1}$ and hence

$$\binom{N}{m_n} \left(\frac{3c}{2} \right)^{m_n} \leq \exp(2^{n-\sigma+1}) \left(\frac{3c}{2} \right)^{m_n} \leq N_{n+\sigma},$$

for sufficiently large σ . Again by Lemma 9.3.12 in [15] we have $\frac{1}{8}m_{n+1} \leq m_n$ for all $n \in \{n_0/2 + \sigma, \dots, n_0 + \sigma - 1\}$ and $m_{n_0+\sigma} = 2^{n_0} \geq N$. It implies that

$$m_n \geq N \left(\frac{1}{8}\right)^{n_0+\sigma-n}. \quad (5.9)$$

Recall that each of the covering cubes has ℓ^2 diameter 2 and therefore by the definition of m_n and then by (5.9)

$$\begin{aligned} e_{n,f}^\tau &\leq \frac{L2^\tau}{\sqrt{N}} \left(d_1(f, 0) \frac{2^n}{\sqrt{m_n}} + d_2(f, 0) 2^{\frac{n}{2}} \right) \\ &\leq L2^{\tau+\frac{n-n_0}{2}} \left(d_1(f, 0) \sqrt{\log(eN/m_n)} + d_2(f, 0) \right) \\ &\leq L2^{\tau+\frac{n-n_0}{2}} (d_1(f, 0) \sqrt{1 + (n_0 + \sigma - n) \log(8)} + d_2(f, 0)). \end{aligned}$$

Again we derive the bound

$$\sum_{n=n_0/2+\sigma}^{n_0+\sigma} e_{n,f}^\tau \leq L2^\tau (d_1(f, 0) + d_2(f, 0)).$$

We have established that

$$\mathcal{E}_{X,f}^\tau(U) = \sum_{n=0}^{\infty} e_{n,f}^\tau \leq L2^\tau (d_1(f, 0) + d_2(f, 0))$$

and consequently

$$\mathcal{E}_{X,\mathcal{F}}^\tau(U) = \sup_{f \in \mathcal{F}} \mathcal{E}_{X,f}^\tau(U) \leq L2^\tau \sup_{f \in \mathcal{F}} (d_1(f, 0) + d_2(f, 0)).$$

By Theorem 2.3 we get

$$\begin{aligned} \mathbf{E} \sup_{f \in \mathcal{F}} \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N |f(X_i)|^2 \right)^{\frac{1}{2}} \\ \leq K \left(\frac{1}{\sqrt{N}} (\gamma_1(\mathcal{F}, d_1) + \gamma_2(\mathcal{F}, d_2)) + \Delta_1(\mathcal{F}) + \Delta_2(\mathcal{F}) \right). \end{aligned}$$

Note that our assumption that d_2 dominates d_1 is not necessary for the result. Clearly using $\bar{d}_2 = \max(d_1, d_2)$ instead of d_2 we only have to observe that for our admissible partition \mathcal{A} which works for $\gamma_1(\mathcal{F}, d_1)$ and $\gamma_2(\mathcal{F}, d_2)$ in the sense of (5.7) one can use the following inequality

$$\sum_{n=0}^{\infty} 2^{\frac{n}{2}} \bar{\Delta}_2(A_n(f)) \leq \sum_{n=0}^{\infty} 2^n \Delta_1(A_n(f)) + \sum_{n=0}^{\infty} 2^{\frac{n}{2}} \Delta_2(A_n(f)),$$

where $\bar{\Delta}_2(A)$ is the diameter of A with respect to \bar{d}_2 distance. In the same way $\bar{\Delta}_2(\mathcal{F}) \leq \Delta_1(\mathcal{F}) + \Delta_2(\mathcal{F})$. We have proved the following result.

Theorem 5.2 *Suppose that $0 \in \mathcal{F}$ and \mathcal{F} satisfies (5.3). Then*

$$\begin{aligned} \mathbf{E} \sup_{f \in \mathcal{F}} \frac{1}{\sqrt{N}} \left(\sum_{i=1}^N |f(X_i)|^2 \right)^{\frac{1}{2}} \\ \leq K \left(\frac{1}{\sqrt{N}} (\gamma_1(\mathcal{F}, d_1) + \gamma_2(\mathcal{F}, d_2)) + \Delta_1(\mathcal{F}) + \Delta_2(\mathcal{F}) \right). \end{aligned}$$

The result is due to Mendelson and Paouris [13] (see Theorem 9.3.1 in [15]) and concerns slightly more general situation. The proof we have shown is different and much less technically involved.

References

- [1] W. BEDNORZ AND R. LATALA (2014) On the boundedness of Bernoulli processes, *Ann. Math.*, **180** 1167-1203.
- [2] S. DILWORTH (1993) The distribution of vector-valued Rademacher series, *Ann. Probab.*, **21**, 2046-2052.
- [3] X. FERNIQUE (1992) Sur la regularite des fonctions aleatoires D'Ornstein-Uhlenbeck valeurs dans l_p , $p \in [1, \infty)$, *Ann. Probab.*, **20**, 1441-1449.
- [4] X. FERNIQUE (1997) Fonctions aléatoires Gaussiennes, vecteurs aléatoires Gaussiens, Université de Montréal, Centre de Recherches Mathématiques, Montreal QC.
- [5] L. KRAWCZYK (1996) Maximal inequality for Gaussian vectors, *Bull. Acad. Polon. Sci. Math.*, **44**, 157-160.
- [6] R.LATALA AND K. OLESZKIEWICZ (1994) On the best constant in the Khintchin-Kahane inequality. *Studia. Math.*, **109**, 101-104.
- [7] R. LATALA (1999) A note on the maximal inequality for VC-classes, *Advances in stochastic inequalities* (Atlanta, GA, 1997), 125-134, Contemp. Math. **234**, Amer. Math. Soc., Providence.
- [8] R.LATALA (2012) Moments of unconditional logarithmically concave vectors, in Geometric Aspects of Functional Analysis, Israel Seminar 2006-2010, 301-315, Lecture Notes in Math. **2050**, Springer 2012.
- [9] R.LATALA (2014) Sudakov-type minoration for log-concave vectors, *Studia Math.* **223**, 251-274.
- [10] R. LATALA AND T. TKOCZ (2015) A note on suprema of canonical processes based on random variables with regular moments, *Electron. J. Probab.* **20**, 1-17.
- [11] M. LEDOUX AND M. TALAGRAND (1991). Probability in Banach spaces. Isoperimetry and processes. *Results in Math. and Rel. Areas (3)*. **23**, Springer-Verlag, Berlin.
- [12] M. B. MARCUS AND G. PISIER (1981) Random Fourier Series with Applications to Harmonic Analysis, *Ann. of Math. Stud.*, **101**, Princeton Univ. Press and Univ. Tokyo Press, Princeton, NJ and Tokyo.
- [13] S. MENDELSON AND G. PAOURIS (2012) On generic chaining and the smallest singular value of random matrices with heavy tails. *J. Funct. Anal.* **262**, 3775-3811.
- [14] S. MENDELSON AND R. VERSHYNIN (2003) Entropy and the combinatorial dimension. *Invent. Math.* **152**, 37-55.
- [15] M. TALAGRAND (2014), Upper and Lower Bounds for Stochastic Processes. Modern Methods and Classical Problems. *A Series of Modern Surveys in Mathematics*, **60**, Springer-Verlag, New-York.